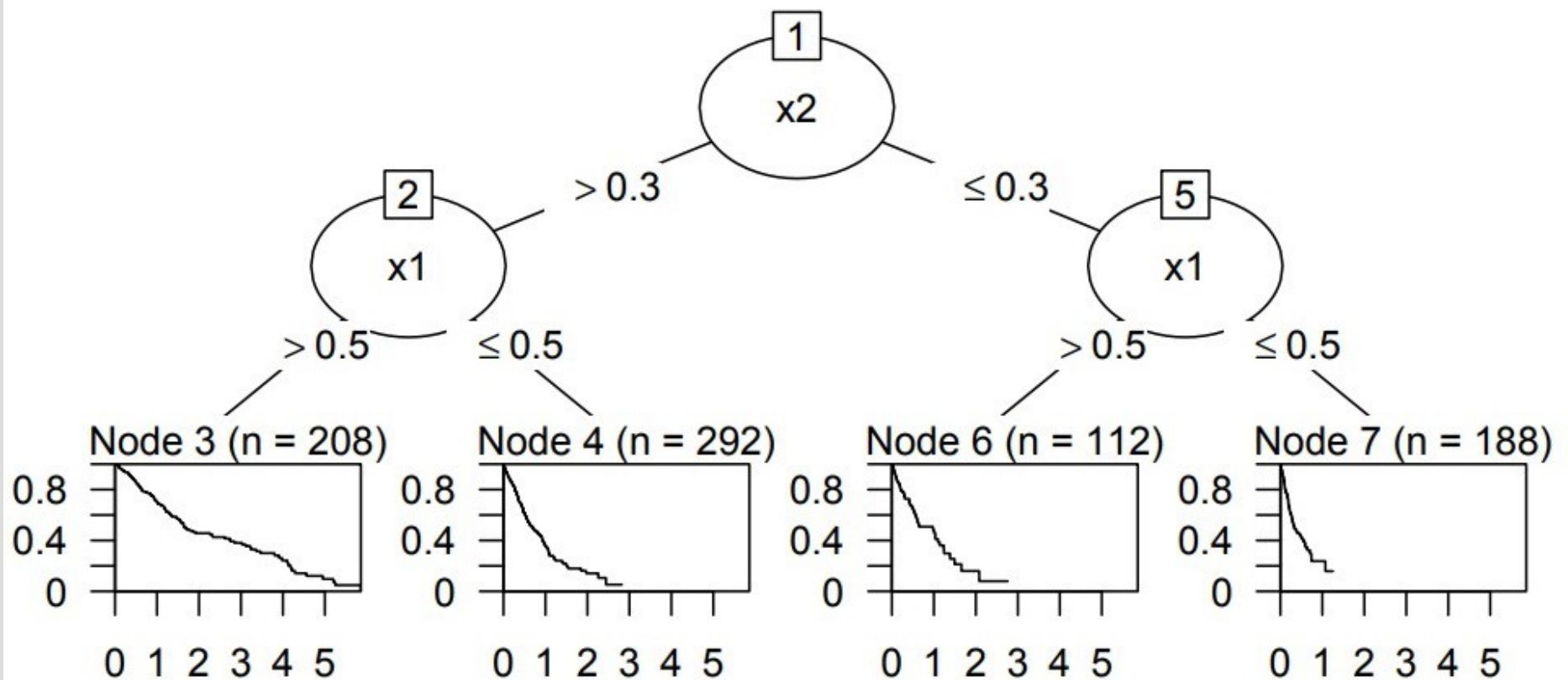


PAKIET MST – Multivariate Survival Trees

Konstruowanie wielowymiarowych drzew przeżycia

Joanna Goławska

Czym są drzewa MST?



Modelowanie przeżycia – czasu do wystąpienia zdarzenia

- Obserwujemy obiekty naturalnie pogrupowane (np. zęby u pacjentów)
- Wielokrotne czasy zdarzeń (utrata kolejnych zębów u pacjenta)
- Cenzurowanie
- Skorelowane czasy zdarzeń
- Klasyczne metody modelowania:
 - model brzegowy Coxa
 - model podatności
- Problem przy dużej liczbie zmiennych

Procedura tworzenia drzewa MST

MST () :

- 1) Budujemy duże drzewo początkowe
- 2) Przycinamy drzewo początkowe w celu uzyskania zbioru poddrzew
- 3) Wybieramy najlepsze drzewo

Podstawowe argumenty funkcji MST ()

- formula – model przeżycia, np. `Surv(time, status) ~ x1 + x2 | id`
- data – zbiór danych
- method – wybór rodzaju modelowania zależności. Dostępne metody: „marginal”, „gamma.frailty”, „exp.frailty”, „stratified”, „independence”
- selection.method – metoda wyboru najlepszego drzewa. Dostępne: „test.sample”, „bootstrap”
- test – w przypadku wybrania `selection.method = „test.sample”`, próba testowa do wyboru drzewa
- B – liczba prób bootstrap w przypadku gdy wybrano `selection.method = „bootstrap”`

Dodatkowe parametry w funkcji `MST()`

- `minsplit` – minimalna liczba obserwacji w węźle do kontynuowania podziału
- `minevents` – minimalna liczba obserwacji niecenzurowanych w węźle do kontynuowania podziału
- `minbucket` – minimalna liczba obserwacji w liściu
- `maxdepth` – maksymalna głębokość drzewa
- `mtry` – liczba zmiennych rozważana przy podziale
- `distinct` – czy rozważane są wszystkie możliwe punkty odcięcia (czy tylko percentyle)
- `sortTrees` – czy drzewa mają być posortowane tak, aby węzeł z lewej strony miał mniejsze ryzyko wystąpienia zdarzenia
- `plot.GA` – czy ma być rysowany wykres oceny drzewa vs rozmiar drzewa

Co zwraca funkcja `MST()` ?

- `tree0` – drzewo początkowe, obiekt `constparty`
- `pruning.info` – przycięte drzewa rozważane przy wyborze najlepszego drzewa
- `best.tree.size` – najlepszy rozmiar drzewa w oparciu o wybraną karę
- `best.tree.structure` – struktura drzew o najlepszym rozmiarze, obiekt `constparty`

Obiekt `constparty` wymaga stałego dopasowania dla każdego liścia. W związku z tym predykcja i rysowanie wykresów ignorują zależności. Zalecane jest, by do predykcji dopasować własny model. (Przykład podany w pliku `MST.R`)

Podsumowanie

- Jedyne dostępne w R pakiety do tworzenia drzew przeżycia
- Pozwala na dużą kontrolę przy konstrukcji drzew
- Drzewa doskonale sobie radzą z danymi zawierającymi ogromną liczbę predyktorów różnych typów nie wymagając przy tym transformacji zmiennych
- Istotny jest wybór modelu – model podatności gamma daje dość dobre wyniki nawet przy złej specyfikacji modelu, jednak w porównaniu z innymi dostępnymi modelami ma znacznie dłuższy czas wykonania
- Dużo dostępnych opcji, które mogą przyspieszyć tworzenie drzewa

Literatura

- „*Constructing Multivariate Survival Trees: The MST Package for R*”, Peter Calhoun, Xiaogang Su, Martha Nunn, Juanjuan Fan, Journal of Statistical Software v. 83, n.12, 2018
- Instrukcja użytkowania pakietu MST,
<https://cran.r-project.org/web/packages/MST/MST.pdf>